

## VU Research Portal

### Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression

Pinto Morais de Carvalho, B.; Postma, C.; Mongera, S.; Hopmans, E.; Diskin, S.; van de Wiel, M.A.; van Crieckinge, W.; Thas, O.; Matthäi, A.; Cuesta valentin, M.A.; Terhaar sive Droste, J.S.; Craanen, M.; Schröck, E.; Ijlstra, B.

#### **published in**

Gut

2009

#### **DOI (link to publisher)**

[10.1136/gut.2007.143065](https://doi.org/10.1136/gut.2007.143065)

#### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

#### **citation for published version (APA)**

Pinto Morais de Carvalho, B., Postma, C., Mongera, S., Hopmans, E., Diskin, S., van de Wiel, M. A., van Crieckinge, W., Thas, O., Matthäi, A., Cuesta valentin, M. A., Terhaar sive Droste, J. S., Craanen, M., Schröck, E., & Ijlstra, B. (2009). Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. *Gut*, 58, 79-89. <https://doi.org/10.1136/gut.2007.143065>

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

#### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

# Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression

B Carvalho,<sup>1</sup> C Postma,<sup>1</sup> S Mongera,<sup>1</sup> E Hopmans,<sup>1</sup> S Diskin,<sup>2,3,4</sup> M A van de Wiel,<sup>1,5,6</sup> W van Criekinge,<sup>7</sup> O Thas,<sup>8</sup> A Matthäi,<sup>9</sup> M A Cuesta,<sup>10</sup> J S Terhaar sive Droste,<sup>11</sup> M Craanen,<sup>11</sup> E Schröck,<sup>9</sup> B Ylstra,<sup>1</sup> G A Meijer<sup>1</sup>

► Additional tables and a figure are published online only at <http://gut.bmj.com/content/vol58/issue1>

<sup>1</sup> Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands; <sup>2</sup> Division of Oncology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; <sup>3</sup> Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA; <sup>4</sup> Penn Center for Bioinformatics (PCBI), University of Pennsylvania; <sup>5</sup> Department of Biostatistics, VU University Medical Center, Amsterdam, The Netherlands; <sup>6</sup> Department of Mathematics, Free University (VU), Amsterdam, The Netherlands; <sup>7</sup> Laboratory of Bioinformatics and Computational Genomics, Department of Molecular Biotechnology, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium; <sup>8</sup> Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Ghent, Belgium; <sup>9</sup> Technical University Dresden, Medical Faculty Carl Gustav Carus, Institute for Clinical Genetics, Dresden, Germany; <sup>10</sup> Department of Surgery, VU University Medical Center, Amsterdam, The Netherlands; <sup>11</sup> Department of Gastroenterology, VU University Medical Center, Amsterdam, The Netherlands

Correspondence to: Dr G A Meijer, Department of Pathology, VU University Medical Center, De Boelelaan 1117, 1081HV Amsterdam, The Netherlands; [ga.meijer@vumc.nl](mailto:ga.meijer@vumc.nl)

Revised 29 July 2008  
Accepted 5 August 2008  
Accepted for Online First 1 October 2008

## ABSTRACT

**Objective:** This study aimed to identify the oncogenes at 20q involved in colorectal adenoma to carcinoma progression by measuring the effect of 20q gain on mRNA expression of genes in this amplicon.

**Methods:** Segmentation of DNA copy number changes on 20q was performed by array CGH (comparative genomic hybridisation) in 34 non-progressed colorectal adenomas, 41 progressed adenomas (ie, adenomas that present a focus of cancer) and 33 adenocarcinomas. Moreover, a robust analysis of altered expression of genes in these segments was performed by microarray analysis in 37 adenomas and 31 adenocarcinomas. Protein expression was evaluated by immunohistochemistry on tissue microarrays.

**Results:** The genes *C20orf24*, *AURKA*, *RNPC1*, *TH1L*, *ADRM1*, *C20orf20* and *TCFL5*, mapping at 20q, were significantly overexpressed in carcinomas compared with adenomas as a consequence of copy number gain of 20q.

**Conclusion:** This approach revealed *C20orf24*, *AURKA*, *RNPC1*, *TH1L*, *ADRM1*, *C20orf20* and *TCFL5* genes to be important in chromosomal instability-related adenoma to carcinoma progression. These genes therefore may serve as highly specific biomarkers for colorectal cancer with potential clinical applications.

The majority of cancers are epithelial in origin and arise through a stepwise progression from normal cells, through dysplasia, into malignant cells that invade surrounding tissues and have metastatic potential. The colorectal adenoma to carcinoma progression is a classic example of this process.<sup>1 2</sup>

Genomic instability is a crucial step in this progression and occurs in two ways in colorectal cancer (CRC).<sup>3</sup> DNA mismatch repair deficiency leading to microsatellite instability (MIN) explains about 15% of cases.<sup>4-6</sup> In the other 85%, genomic instability occurs at the chromosomal level (CIN) giving rise to aneuploidy. While for a long time chromosomal aberrations have been regarded as random noise, it is now well established that these DNA copy number changes occur in specific patterns and are associated with different clinical behaviour.<sup>7-9</sup> Nevertheless, neither the cause of CIN in human cancer progression nor its biological consequences have been fully appreciated.

Chromosomal aberrations frequently reported in CRC are 7p, 8q, 13q and 20q gains, and 4p, 5q, 8p, 15q, 17p and 18q losses.<sup>10-13</sup> Of these, especially 8q, 13q and 20q gains and 8p, 15q, 17p and 18q

losses are associated with colorectal adenoma to carcinoma progression.

Gain of 20q is observed in >65% of CRCs.<sup>14</sup> Gains of 20q are also common in other tumour types and have been associated with poor outcome in gastric cancer and CRC.<sup>15-20</sup> The 20q13 amplicon has been studied in detail in breast and gastric cancers with restricted contig array comparative genomic hybridisation (CGH), pinpointing several genes as targets of amplification.<sup>21 22</sup> Analysis of DNA copy number changes at the gene level by multiplex ligation-dependent probe amplification showed that in CRC, besides 20q13, 20q11 is also frequently amplified.<sup>23</sup>

This study aims to investigate dosage effects of putative 20q oncogenes in colorectal adenoma to carcinoma progression.

## MATERIALS AND METHODS

### Tumour samples

Forty-one formalin-fixed and paraffin-embedded progressed adenomas (with a focus of carcinoma present, also referred to as malignant polyps) collected from the tissue archive of the department of pathology at the VU University Medical Center (VUmc), Amsterdam, The Netherlands, and 73 prospectively collected snap-frozen colorectal tumour samples (37 non-progressed adenomas and 36 carcinomas) were investigated. All samples were used in compliance with the institution's ethical regulations.

The 41 progressed adenomas corresponded to 19 females and 18 males (three patients presented more than one lesion). Mean age was 67 years (range 45–86). From these, adenoma and carcinoma components were analysed separately, adding up to a total of 82 archival samples (41×2).

The 73 frozen specimens corresponded to 31 females and 34 males (six patients had multiple tumours). Mean age was 69 years (range 47–89). All histological sections were evaluated by a pathologist (G.A.M.).

Array CGH was performed on both sets of samples, while expression microarrays were performed on the frozen samples only.

### DNA and RNA isolation

DNA from paraffin-embedded samples was obtained as described previously.<sup>24</sup> RNA and DNA from snap-frozen tissues were isolated using TRIzol (Invitrogen, Breda, The Netherlands)

## Colorectal neoplasia

following the supplier's instructions with some modifications, described at <http://www.english.vumc.nl/afdelingen/microarrays>. Isolated RNA was subjected to purification using an RNeasy Mini Kit (Qiagen, Venlo, The Netherlands). RNA and DNA concentrations and purities were measured on a Nanodrop ND-1000 spectrophotometer (Isogen, IJsselstein, The Netherlands), and integrity was evaluated on a 1% agarose ethidium bromide-stained gel.

### Array CGH

A BAC/PAC (bacterial artificial chromosome/phage artificial chromosome) array platform was used as described elsewhere.<sup>25</sup> Arrays were scanned (Agilent DNA Microarray scanner G2505B, Agilent Technologies, Palo Alto, California, USA) and Image 5.6 software (Biodiscovery, Marina del Rey, California, USA) was used for automatic feature extraction with default settings. Local background was subtracted from the signal median intensities of both test and reference DNA. The median of the triplicate spots was calculated for each BAC clone, and  $\log_2$  ratios (tumour/normal) were normalised by subtraction of the mode value of BAC clones on chromosomes 1–22 (UCSC July 2003 freeze of the Human Golden Path-NCBI Build 34). Clones with standard deviation of the intensity of the three spots  $> 0.2$  and with  $>20\%$  missing values were excluded.

### Expression microarrays

The Human Release 2.0 oligonucleotide library, containing 60mer oligonucleotides representing 28 830 unique genes, designed by Compugen (San Jose, California, USA) was obtained from Sigma-Genosys (Zwijndrecht, The Netherlands). Printing of slides was done as described elsewhere.<sup>26</sup> Tumour RNA (30  $\mu\text{g}$ ) was hybridised against Universal Human reference (Stratagene, Amsterdam, The Netherlands). cDNA labelling and hybridisation procedures are described elsewhere.<sup>26</sup> Scanning of arrays and feature extraction were performed as described above. Overall quality of experiments was judged on microarray plots of intensities of raw data. Normalisation was done either with TIGR Midas (<http://www.tm4.org/midas.html>), using "Lowess" correction<sup>27</sup> or with "Median" normalisation and implemented in the maNorm function (Marray R bioconductor package), with identical results. Interarray normalisation was also performed. Low intensity values were replaced by the intensity value of 50. Genes with  $>20\%$  missing values were excluded.

Array CGH and expression microarray data sets are available at Gene Expression Omnibus (GEO) <http://www.ncbi.nlm.nih.gov/geo/>,<sup>28</sup> accession number GSE8067.

### Microarray data analysis

Below, the steps of data analysis are discussed for array CGH data, expression data and integrative analysis. To account for multiple testing, either a false discovery rate (FDR) correction was applied to the p values, or a very stringent p value cut-off was used.

### Array CGH data

To segment DNA copy number alterations, a smoothing algorithm, "aCGH-Smooth", was applied.<sup>29</sup> Smoothed  $\log_2$  ratios of  $-0.15$  and  $0.15$  were used as thresholds to define gains and losses (99% CIs) obtained for 15 normal-to-normal hybridisations. Only gains and losses covering at least three consecutive BAC clones were included. Amplifications were called when  $\log_2$  ratios exceeded 1.0. DNA copy number data

were stored in ArrayCGHbase<sup>30</sup> (<http://arraydb.vumc.nl/arrayCGHbase>). Median absolute deviation (MAD) was determined for each case as a quality control. Cases with  $\text{MAD} \geq 0.2$  were excluded. Array CGH profiles were visualised in ArrayCGHbase.

Supervised analysis, comparing two groups, was done using CGHMultiArray.<sup>31</sup> For analysis of paired samples (adenoma and carcinoma components within progressed adenomas), an adapted version of CGHMultiArray was used, based on the Wilcoxon sign-rank test corrected for ties. Reported p values are adjusted for multiple testing (FDR), unless stated otherwise.

For defining the most frequent smallest regions of overlap (SRO) for gains on 20q, in all cases, STAC (significance testing for aberrant copy number) was used.<sup>32</sup>

### Microarray expression data

As all hybridisations were performed against a common reference, all comparisons were relative between colorectal adenomas and carcinomas.

Supervised analysis for comparing carcinomas and adenomas was done using the Wilcoxon signed rank test, and a modified version of this test—the total Thas score (<http://www.cvstat.ugent.be/index.php?page=techrep/techrep.htm>) that is powerful when the distributions of the expression levels of both groups do not differ over the whole range of expression levels. This occurs when not all cases in the carcinoma and adenoma groups have differentially expressed genes, but differences rather appear in subpopulations. Genes were considered as differentially expressed when a Wilcoxon test p value was  $< 1e-5$  and a Thas p value was  $< 0.05$ , corresponding to an FDR of  $< 0.05$ .

To disclose genes whose expression is influenced by 20q gain, tumours with and without 20q gain were compared. Gene expression was regressed on copy number count using a linear model.

To evaluate the discriminatory power of candidate genes for classifying adenomas versus carcinomas, a stepwise linear discriminant analysis with leave one out cross-validation was performed on mRNA expression data (SPSS 15.0 for Windows, SPSS, Chicago, Illinois, USA).

### Integration of copy number and expression data

ACE-it (Array CGH Expression integration tool) was applied to test whether gene dosage affects RNA expression.<sup>33</sup> Only genes on chromosome 20 are presented. We used a cut-off value of 0.15 for gains and losses, a default group value of 9 and an FDR  $\leq 0.10$ .

### Quantitative reverse transcription-PCR (qRT-PCR)

RNA (1  $\mu\text{g}$ ) was treated with DNase I and reverse transcribed to cDNA using oligo(dT)<sub>20</sub> primer with Superscript II reverse transcriptase (Invitrogen).

qRT-PCR was performed in duplicate on 15 adenomas and 15 carcinomas for six candidate genes. A master mix was prepared with 12.5  $\mu\text{l}$  of SYBR Green PCR master mix (Applied Biosystems, Nieuwerkerk a/d IJssel, The Netherlands), 0.5  $\mu\text{M}$  of each primer in 22.5  $\mu\text{l}$ . cDNA (25 ng in 2.5  $\mu\text{l}$ ) was added to the mix. Reactions were performed in a 7300 Real-time PCR System (Applied Biosystems). Amplification conditions comprised a denaturation step at 95°C for 10 min and 50 cycles at 95°C for 15 s and annealing temperature for 1 min (Supplementary table 1). Relative expression levels were determined following the  $2^{-\Delta\Delta\text{Ct}}$  method,<sup>34</sup> using the  $\beta$ -2-microglobulin gene as a reference. This gene was previously

**Table 1** Genes significantly upregulated in carcinomas, when compared with adenomas, mapping at 20q (Wilcoxon ranking p value <1e-5 (ie, FDR<0.05) and Thas p value <0.05), ordered by chromosomal position (location in base pairs (bp) according to freeze July 2003; NCBI Build 34) with HUGO gene symbols and GenBank accession ID

Gene symbol	GenBank accession no.	Location (bp position)	Wilcoxon p value	Thas p value
<i>C20orf1 (TPX2)</i>	NM_012112	31 103 374	2E-06	8E-05
<i>MYRL2</i>	NM_006097	35 859 501	5E-06	4E-05
<i>C20orf24 (RIP5)</i>	NM_018840	35 923 014	2E-07	2E-05
<i>TOMM34</i>	NM_006809	44 265 329	8E-08	0
<i>RBPSUHL</i>	NM_014276	44 626 010	2E-07	6E-06
<i>BCAS4</i>	NM_017843	50 138 063	2E-06	6E-05
<i>AURKA (STK6)</i>	NM_003600	55 641 283	4E-10	0
<i>FLJ37465 (BMP7)</i>	AK094784	56 477 906	1E-09	0
<i>RNPC1</i>	NM_017495	56 660 843	8E-07	7E-05
<i>TH1L</i>	NM_016397	58 253 070	1E-06	1E-05
<i>ADRM1</i>	NM_007002	61 566 389	9E-07	8E-05
<i>C20orf20</i>	NM_018270	62 156 238	9E-09	0
<i>TCFL5</i>	NM_006602	62 211 152	2E-09	0
<i>C20orf11</i>	NM_017896	62 299 593	4E-07	0

FDR, false discovery rate.

demonstrated not to differ in expression between adenomas and carcinomas.<sup>35</sup>

### Immunohistochemistry on tissue microarrays (TMAs)

A TMA was constructed with 57 tumours (32 adenomas and 25 carcinomas) for which array CGH and/or expression microarray data were available. For each tumour, three cores from different locations within the tumour were included in the array. A 4 µm section of the array was used for immunohistochemistry. After deparaffination in xylene, and rehydration through graded alcohol to water, endogenous peroxidase was blocked with hydrogen peroxide (0.3% H<sub>2</sub>O<sub>2</sub>/methanol) for 25 min. Antigen retrieval was done by autoclaving in citrate buffer (10 mM; pH 6.0). Primary Aurora A monoclonal antibody NCL-L-AK2 from Novocastra Laboratories (Newcastle, UK) was incubated overnight at 4°C at a dilution of 1:50. The secondary mouse antibody, K4006, from the Envision kit (DAKO, Heverlee, Belgium) was incubated for 30 min at room temperature. Counterstaining was done with Mayer's haematoxylin. Incubation without primary antibody was used as negative control. The colorectal cancer cell line Caco-2, which has a 20q gain and is known to express Aurora A, was used as positive control. Caco-2 cells were fixed and paraffin embedded, and sections were taken along in the same run of immunohistochemistry as that in which the TMA was processed. Caco-2 produced strong nuclear, mostly along with cytoplasmic, staining in >75% of tumour cells, and this pattern was taken as reference for intense staining. Next, the spectrum of staining in the respective cores on the TMA was surveyed in terms of intensity and positive nuclei. Only staining in tumour cells (ie, either adenoma or carcinoma cells) was considered. Cores of the TMA typically contained 4–17 crypts with every crypt >100 cells, which all were evaluated. Basically, three staining patterns were seen: no staining at all, strong staining comparable with that observed in Caco-2 cells; and an intermediate pattern that showed positive staining, but clearly less intense than in Caco-2 cells. The intensity of staining was taken as the most important parameter. In pattern 2, typically 50% to >75% of nuclei showed intense staining, while in pattern 1 typically 25% to >75% of nuclei showed weak staining. For score 0, no more

than a scattered weakly positive cell was tolerated. Based on evaluation of up to three cores by two independent observers, a score ranging from 0 to 2 was assigned per tumour, with score 0 corresponding to no signal, score 2 corresponding to the strong signal that was observed in the positive control Caco-2, and score 1 for an intermediate intensity staining. In the case of disagreement between observers, a third observer was consulted and the majority score was noted.

Cochran–Armitage test analysis was performed to compare protein expression with lesion type (adenoma, carcinoma). The Jonckheere–Terpstra test was performed to compare protein expression with log<sub>2</sub> ratios (microarray expression data). Both tests make explicit use of the ordinality of the protein levels of expression. Differences were considered significant when p <0.05.

## RESULTS

### Delimiting gained regions on 20q

We analysed by array CGH 41 progressed adenomas, which were previously studied by classical CGH. We analysed the adenoma and carcinoma components of these samples separately. Gain of 20q occurred in >60% of the cases (fig 1A,B; Supplementary fig 1A). The pattern of copy number changes did not differ between adenoma and carcinoma components (as determined by CGHMultiArray), although it sometimes showed lower amplitudes in the adenoma component (fig 1A,B).

Next, we analysed the DNA copy number status of 37 non-progressed adenomas and 36 carcinomas. From these 73 tumours, 67 (34 adenomas and 33 carcinomas) showed high quality genomic profiles with MAD values <0.2, giving an 8% drop-out. In these 67 tumours, chromosome 20 gain occurred in <15% of the adenomas but in >60% of the carcinomas (p <0.001, as determined by CGHMultiArray), mostly affecting either all of chromosome 20 or the q-arm only, similar to the progressed adenomas (fig 1C,D; Supplementary fig 1B).

To determine the most relevant regions within 20q harbouring putative oncogenes with a role in colorectal adenoma to carcinoma progression, STAC<sup>32</sup> was applied to the combined set of paraffin-embedded malignant polyps (n=41) and frozen carcinomas (n=33). This revealed three relevant regions of aberrant copy gains on 20q, one spanning 4 Mb (32–36 Mb), one spanning 3 Mb (56–59 Mb) and the third one spanning 2 Mb (61–64 Mb) (fig 2). These three regions (SROs) contained 80, 35 and 94 known genes, respectively.

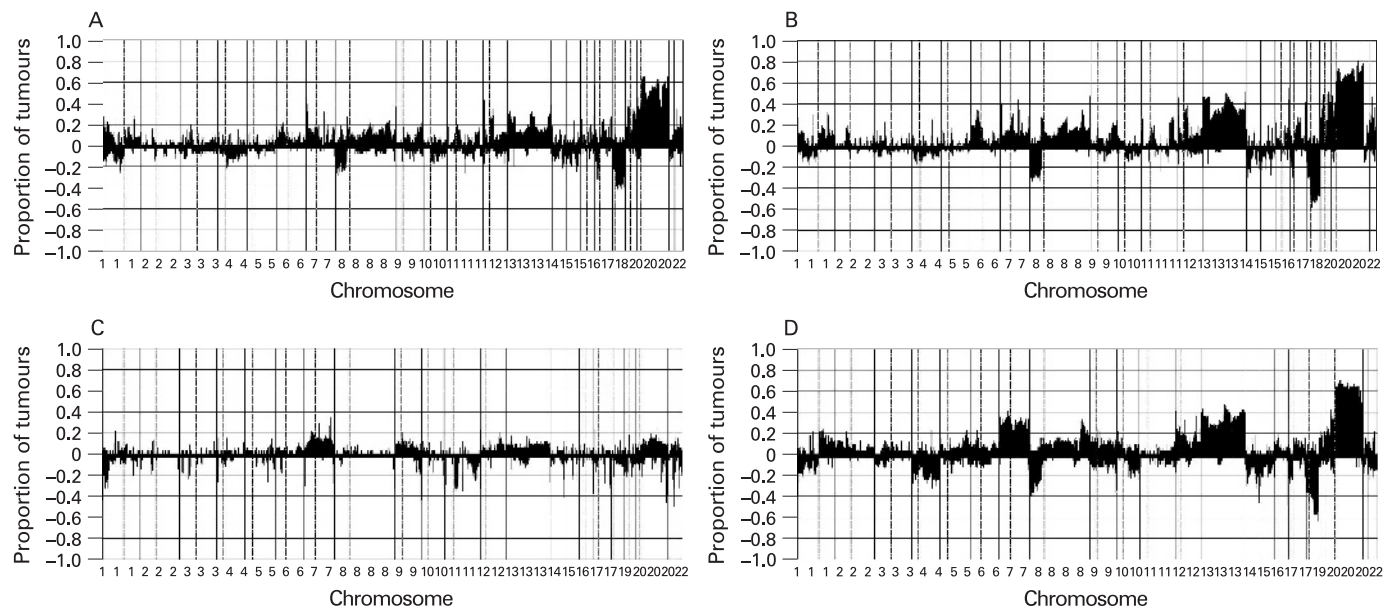
### Identification of differentially expressed genes

We performed microarray expression analysis on the 37 non-progressed adenomas and 36 carcinomas for which snap-frozen material was available. High quality expression array data were obtained from 68 cases (37 adenomas and 31 carcinomas, 7% drop-out).

Supervised data analysis for identifying putative oncogenes on 20q was done in two different ways: we compared carcinomas with adenomas, and we compared tumours with 20q gain with tumours without 20q gain. The first approach revealed genome-wide 122 upregulated genes and 219 down-regulated genes (a total of 341 differentially expressed genes) in carcinomas when compared with adenomas (Wilcoxon test p value <1e-5 (FDR<0.05) and Thas p value <0.05). Of these 122 upregulated genes, 14 map at chromosome 20q (table 1). For the second approach, only tumours (adenomas and carcinomas) that had both array CGH data and expression data available (n=64) were included. As a preselection, we used genes



## Colorectal neoplasia



**Figure 1** Frequency plot of DNA copy number gains and losses determined by bacterial artificial chromosome (BAC) array comparative genomic hybridisation in (A) adenoma components of 41 progressed colorectal adenomas, (B) adenocarcinoma components of 41 progressed colorectal adenomas, (C) 34 non-progressed colorectal adenomas and (D) 33 adenocarcinomas. The y-axis displays the fraction of tumours with either a gain (positive sign) or loss (negative sign) for all clones that are sorted by chromosome and base pair position.

differentially expressed (both upregulated and downregulated) between carcinomas and adenomas, as we focus on genes at 20q that are involved in progression, using a less stringent cut-off (Thas  $p$  value  $<0.05$ ). Thereby, we identified 127 genes genome-wide out of 931 differentially expressed genes (regression analysis; FDR  $\leq 0.1$ ) whose expression levels are influenced by the occurrence of 20q gain. Of these 127 genes, 21 are mapped at 20q (table 2).

Nine genes common to these two approaches emerged, namely *TPX2*, *C20orf24*, *AURKA*, *RNPC1*, *TH1L*, *ADRM1*, *C20orf20*, *TCFL5* and *C20orf11*.

#### Integration of array CGH and expression data

BAC array CGH data were related to oligonucleotide expression array data, independently of adenoma or carcinoma status, using a dedicated integration tool called ACEit.<sup>33</sup> We obtained a list of 151 genes located at chromosome 20 for which gene dosage affected expression levels (FDR  $\leq 0.1$ ), 120 of which are on the q-arm (Supplementary table 2). Combining this information with the results of the two supervised approaches for expression data analysis (carcinoma vs adenoma and 20q gain vs no 20q gain), seven genes were shared (fig 3). For these genes, *C20orf24*, *AURKA*, *RNPC1*, *TH1L*, *ADRM1*, *C20orf20* and *TCFL5*, combined box plots with dot plots of mRNA expression in adenomas versus carcinomas (fig 4A) and scatter plots of mRNA expression versus DNA copy number ratio (fig 4B) are shown.

Of these seven candidate genes, six map within the SROs determined by STAC analysis. The seventh gene (*AURKA*) maps approximately 400 kb proximal to SRO2 at 55.6 Mb (20q13.31). *C20orf24* maps within SRO1 at 35.9 Mb (20q11.23), *RNPC1* and *TH1L* map within SRO2 at position 56.7 and 58.3 Mb, respectively (20q13.32), and *ADRM1*, *C20orf20* and *TCFL5* map within SRO3, the first at 61.6 Mb and the other two at 62.2 Mb (20q13.33).

Stepwise linear discriminant analysis with leave one out cross-validation showed that mRNA expression levels of two out of the seven candidate genes—that is, *RNPC1* and *TCFL5*—allowed

88.2% of the cases (60/68) to be correctly classified as adenomas or carcinomas (fig 5 and table 3).

#### Confirmation of differential expression by qRT-PCR and immunohistochemistry

qRT-PCR was performed on a subsample ( $n=30$ ) of frozen tumours (15 adenomas and 15 carcinomas) to confirm the expression levels of six of the seven genes identified.

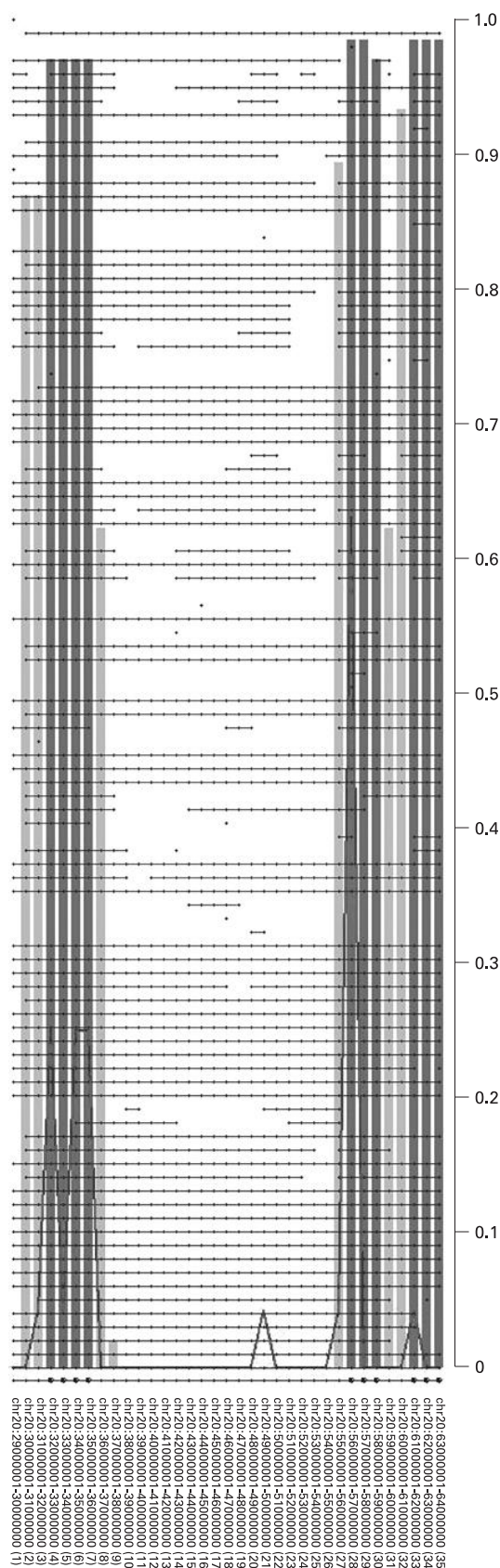
Carcinomas showed higher expression of all six genes compared with adenomas, and tumours with 20q gain (4 adenomas and 8 carcinomas) showed higher expression compared with tumours without 20q gain (11 adenomas and 7 carcinomas). Table 4 shows the fold changes observed between either carcinomas and adenomas or tumours with 20q gain vs tumours without 20q gain, by microarrays and by qRT-PCR.

In situ confirmation of *AURKA* protein expression by immunohistochemistry on TMAs yielded higher expression of *AURKA* in carcinomas compared with adenomas ( $p=0.01$ ) (table 5) as well as a significant positive correlation with the mRNA expression levels ( $p=0.01$ ) (fig 6). Validation of other genes was hampered by the absence of adequate antibodies.

#### DISCUSSION

One of the most frequent chromosomal aberrations observed in CRC is gain of the long arm of chromosome 20. Nonetheless, which of the many genes mapping at 20q show altered expression due to DNA copy number alterations and play a role in the progression of colorectal adenoma to carcinoma is not yet fully understood. In order to try to identify these putative oncogenes, we analysed a series of colorectal tumours, both adenomas and carcinomas, at the DNA and RNA level.

In this study, we confirmed that chromosome 20 was the most frequently altered in the progressed adenomas and carcinomas (in  $>60\%$  of cases). In non-progressed adenomas, gains of 20q were detected in  $<20\%$ , supporting a role for 20q gain in colorectal adenoma to carcinoma progression consistent with earlier observations.<sup>7</sup> Narrowing down the gained region



**Figure 2** Delimitation of the smallest regions of overlap by STAC (significance testing for aberrant copy number) analysis for 115 samples (41 non-progressed adenomas, 41 adenocarcinoma components of progressed adenomas, and 33 adenocarcinomas). Results for the long arm of chromosome 20 are displayed. Rows represent samples, and columns represent chromosomal locations. A black dot indicates a gain

**Table 2** Genes whose expression is related to the 20q gain (FDR  $\leq 0.10$ ), ordered by chromosomal position (location in base pairs (bp) according to freeze July 2003; NCBI Build 34) with HUGO gene symbols and GenBank accession ID

Gene symbol	GenBank accession no.	Location (bp position)	FDR
<i>HM13</i>	NM_030789	30 874 805	0.03
<i>C20orf1 (TPX2)</i>	NM_012112	31 103 374	0.03
<i>CDC91L1</i>	NM_080476	33 922 394	0.02
<i>C20orf44</i>	NM_018244	34 608 051	0.07
<i>DLGAP4</i>	NM_014902	35 761 669	0.05
<i>TGIF2</i>	NM_021809	35 897 616	0.003
<i>C20orf24 (RIP5)</i>	NM_018840	35 923 014	0.0006
<i>YWHAB</i>	NM_014052	44 210 177	0.0002
<i>UBE2C</i>	NM_007019	45 128 792	0.01
<i>DPM1</i>	NM_003859	50 248 672	0.000001
<i>NFATC2</i>	AK025758	50 769 018	0.003
<i>AURKA (STK6)</i>	NM_003600	55 641 283	0.02
<i>RNPC1</i>	NM_017495	56 660 843	0.04
<i>TH1L</i>	NM_016397	58 253 070	0.007
<i>ADRM1</i>	NM_007002	61 566 389	0.05
<i>SLCO4A1</i>	NM_016354	62 015 102	0.08
<i>C20orf20</i>	NM_018270	62 156 238	0.04
<i>TCFL5</i>	NM_006602	62 211 152	0.03
<i>C20orf11</i>	NM_017896	62 299 593	0.0009
<i>C20orf59</i>	NM_022082	62 323 360	0.007
<i>PRPF6</i>	NM_012469	63 364 789	0.03

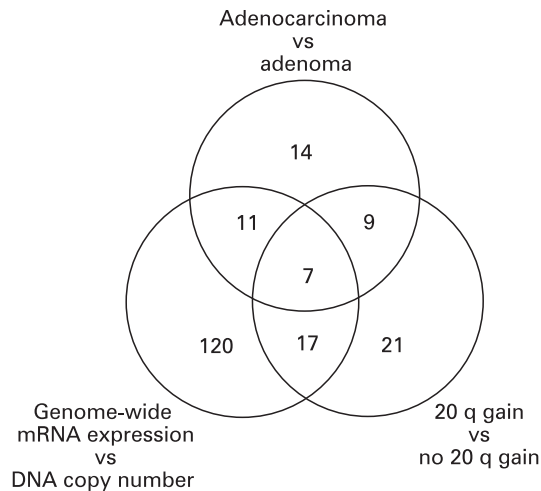
FDR, false discovery rate.

by array CGH across all tumours analysed yielded three SROs: SRO1 at 20q11.22–q11.23 (32–36 Mb), SRO2 at 20q13.32–q13.33 (56–59 Mb) and SRO3 at 20q13.33 (61–64 Mb). Douglas and collaborators,<sup>12</sup> in a series of 48 cell lines and 37 primary CRCs, defined 20q13.3 as the most common region of gain by array CGH (at 62.3 Mb), which is included in SRO3. Moreover, Nakao and collaborators,<sup>13</sup> also by array CGH in a panel of 125 primary CRCs, observed like us more than one prominent region of gain on 20q, centring at 32.3, 37.8, 45.4, 54.7, 59.4 and 65 Mb (according to the August 2001 freeze of the human genome). Based on these array CGH data, the previous studies listed several candidate oncogenes located at these genomic regions, such as *AIB4* (35 Mb), *ZNF217* (53.9 Mb), *CYP24* (54.5 Mb) and *AURKA* (56.6 Mb), and at 62.3 Mb the genes *LIVIN*, *PTK6*, *HD54* and *EEF1A2*. In the present study, the defined regions SRO1, 2 and 3 harbour 80, 35 and 94 genes, respectively, including *AURKA*, *PTK6* and *EEF1A2*. Altogether, based on copy number alterations alone, the number of genes is still too high (209 genes, in total) really to pinpoint the genes that play a role in the progression of colorectal adenoma to carcinoma.

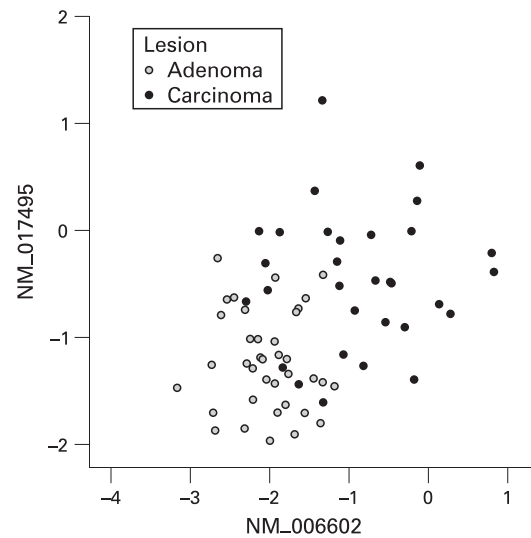
Comparison of the expression of carcinomas relative to adenomas using strict statistical parameters showed that 14 genes mapping at 20q were overexpressed in carcinomas as compared with adenomas. Most of these genes were not described in previous expression microarray studies comparing adenomas with carcinomas.<sup>36 37</sup> These studies used either Affymetrix arrays representing 6600 genes or cDNA microarrays representing 23 000 genes as opposed to the 28 830 represented on our oligonucleotides arrays. In addition, the number of

called in a sample at a location. Consecutive black dots are connected by a line to represent an interval of aberration. Grey bars track the maximum STAC confidence (1 – p value); darker bars are those with confidence >0.95. The line graph indicates the actual frequencies in the sample set.

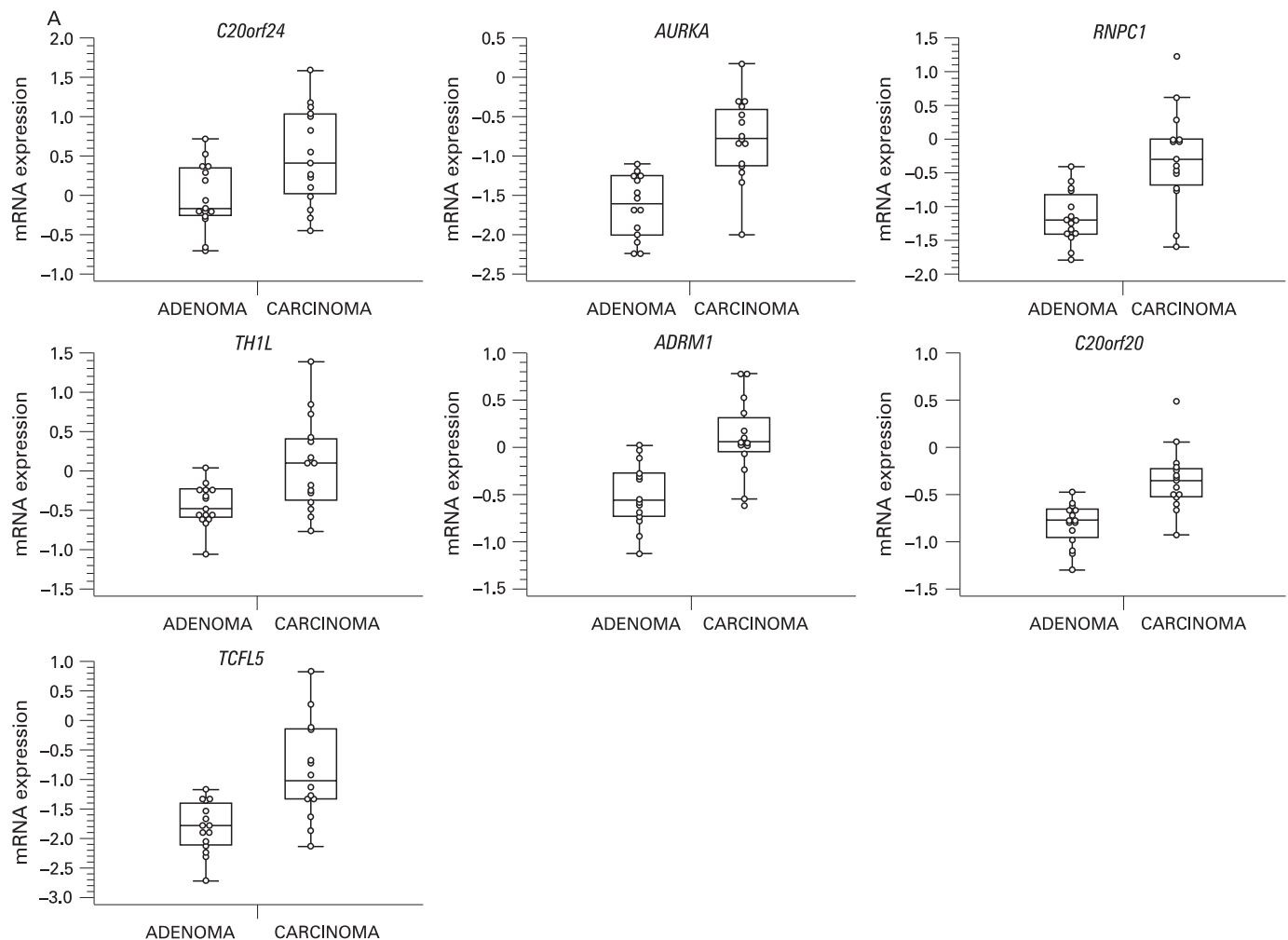
## Colorectal neoplasia

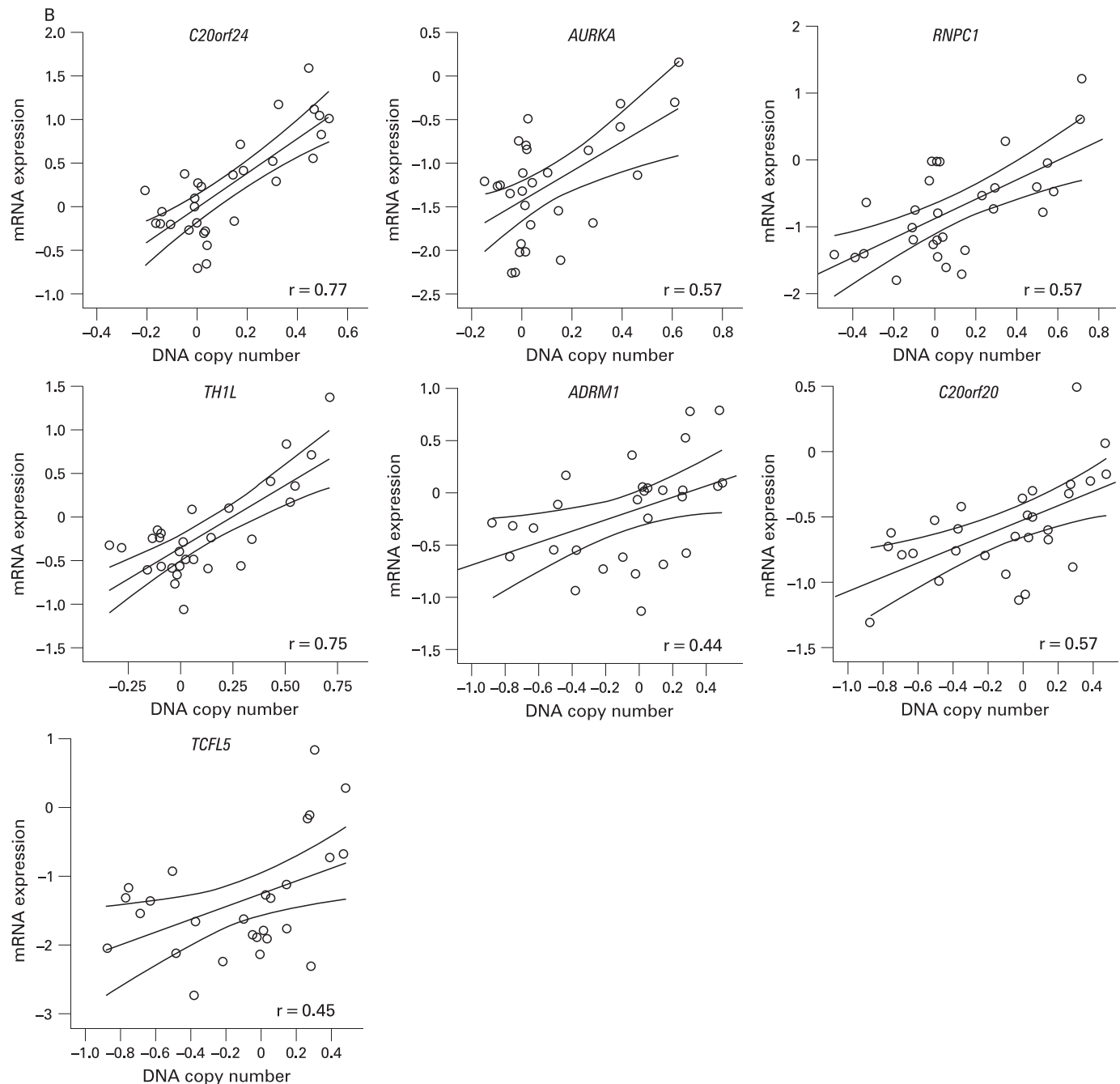


**Figure 3** Venn diagram integrating the results of three different data analysis approaches (comparing colorectal adenocarcinomas vs adenomas; colorectal tumours with 20q gain vs tumours without 20q gain; and genome-wide integration of mRNA expression data with DNA copy number data). Seven genes (*C20orf24*, *AURKA*, *RNPC1*, *TH1L*, *ADRM1*, *C20orf20* and *TCFL5*) emerge with all three approaches.



**Figure 5** Scatter plot of mRNA expression levels of *RNPC1* (NM\_017495) and *TCFL5* (NM\_006602), by lesion (grey circles, adenomas; black circles, carcinomas) showing a good separation of colorectal adenomas vs carcinomas.





**Figure 4** Integration of expression microarray data and array CGH (comparative genomic hybridisation) data of genes *C20orf24*, *AURKA*, *RNPC1*, *TH1L*, *ADRM1*, *C20orf20* and *TCFL5*. (A) Combined box plots with dot plots of mRNA expression (determined by oligonucleotide microarrays) in colorectal adenomas and carcinomas. (B) Scatter plots showing correlation of mRNA expression (determined by oligonucleotide microarrays) and DNA copy number (determined by bacterial artificial chromosome (BAC) array CGH).

samples analysed in the present study is considerably higher, with 4–10 times more adenomas and 2–3 times more carcinomas. In combination with our multiangle, integrative data analysis, the power of the present study thus is higher. Two genes that have been described as being overexpressed in colon tumours, compared with normal mucosa, are *ADRM1* and *TOMM34*.<sup>38</sup> Moreover, cDNA expression analysis in a series of clinical samples of CRC patients showed frequent upregulation of *TOMM34* in carcinomas, and inhibition of this gene by small interfering RNA (siRNA) in the HCT116 colon cancer cell line drastically reduced the cell growth.<sup>39</sup> Together, these findings corroborate our results since *ADRM1* and *TOMM34* are on the

list of genes that were shown to be upregulated in carcinomas (compared with adenomas) (table 1). This lends support to the approach of the present study, and suggests also that the other candidates identified might be relevant.

Looking at the same expression data from a different angle—that is, comparing the expression of tumours with and without 20q gain—we aimed to find genes with a dosage effect on expression. Genome-wide, expression of 127 out of 931 genes was related to 20q gain, 21 of which are located at chromosome 20q itself. Although chromosome 20 has a high gene density, and copy number gains of the long arm are very frequent, certainly not all genes mapping at the gained regions are



## Colorectal neoplasia

**Table 3** Results of stepwise linear discriminant analysis with leave one out cross-validation of the seven candidate genes

Lesion	Predicted group membership		
	Adenoma	Carcinoma	Total
Original Count			
Adenoma	35	2	37
Carcinoma	6	25	31
%			
Adenoma	94.6	5.4	100.0
Carcinoma	19.4	80.6	100.0

From 68 tumours in total, 60 were correctly classified (88.2%), using expression levels of *RNPC1* and *TCFL5* only.

recurrently overexpressed. A total of 209 genes are mapped to the SROs defined here, but only 21 genes are recurrently upregulated in association with 20q gain, consistent with earlier observations.<sup>40–41</sup>

Nine genes overlapped between the 14 adenoma versus carcinoma genes and the 21 genes associated or not with 20q gain, namely *TPX2*, *C20orf24*, *AURKA*, *RNPC1*, *TH1L*, *ADRM1*, *C20orf20*, *TCFL5* and *C20orf11*.

*TPX2* and *AURKA*, differentially expressed between carcinomas and adenomas and associated with 20q gain, are known to be related to aneuploidy and interact with each other.<sup>42</sup> Moreover, in a microarray-based CIN gene signature, *TPX2* ranked first.<sup>43</sup> This finding supports the hypothesis that chromosome 20q gain has an important role in colorectal adenoma to carcinoma progression, and that the onset of manifest CIN is instrumental.

Our third approach, integration of DNA copy number changes and gene expression, demonstrated that throughout the genome 507 genes showed a statistically significant association between DNA copy number and mRNA expression status, for both amplified/upregulated and deleted/down-regulated genes, 120 of these being located on chromosome 20q. Some well-known genes, such as *ZNF217* and *CSE1L*, previously reported to be amplified in colon cancer,<sup>44–46</sup> showed relatively higher expression when there was allelic gain. From these 120 genes, 17 overlapped with the 20q gain-associated list, and 11 overlapped with the adenoma and carcinoma

**Table 5** *AURKA* protein expression in colorectal adenomas vs carcinomas by immunohistochemistry on tissue microarray

Lesion	AURKA staining			Total	p Value*
	Negative	Weak	Strong		
Adenoma	12	12	1	25	
Carcinoma	4	9	6	19	
Total	16	21	7	44	0.01

\*Cochran–Armitage test.

versus carcinoma list. Overlapping these three approaches (expression in adenomas vs carcinomas, expression with no 20q gain vs with 20q gain, and genome-wide expression vs whole-genome copy number changes) showed that seven genes are consistently significant (fig 3), namely *C20orf24*, *AURKA*, *RNPC1*, *TH1L*, *ADRM1*, *C20orf20* and *TCFL5*. In addition to the already stringent data analysis, a permutation analysis was performed, comparing the differential expression of the seven 20q genes with the expression of >50 000 random subsets out of 7946 genes in silent DNA regions (2q, 3, 5, 10p, 11, 16, 21 and 22). For each random subset, the Wilcoxon scores of the seven most differentially expressed (adenoma vs carcinoma) genes were selected. The seven genes on 20q showed a significantly higher expression in adenocarcinomas versus adenomas compared with the best performing combination from the permutation test ( $p = 0.001$ ), underlining that the copy number-based discovery of putative oncogenes did not yield random differentially expressed genes. The fact these overexpressed putative oncogenes on 20q actually resulted in biologically active components (ie, proteins) in the tumour cells was demonstrated by immunohistochemistry on TMA for *AURKA*. For the other candidates, antibodies did not perform adequately in the tissue samples or were not available at all.

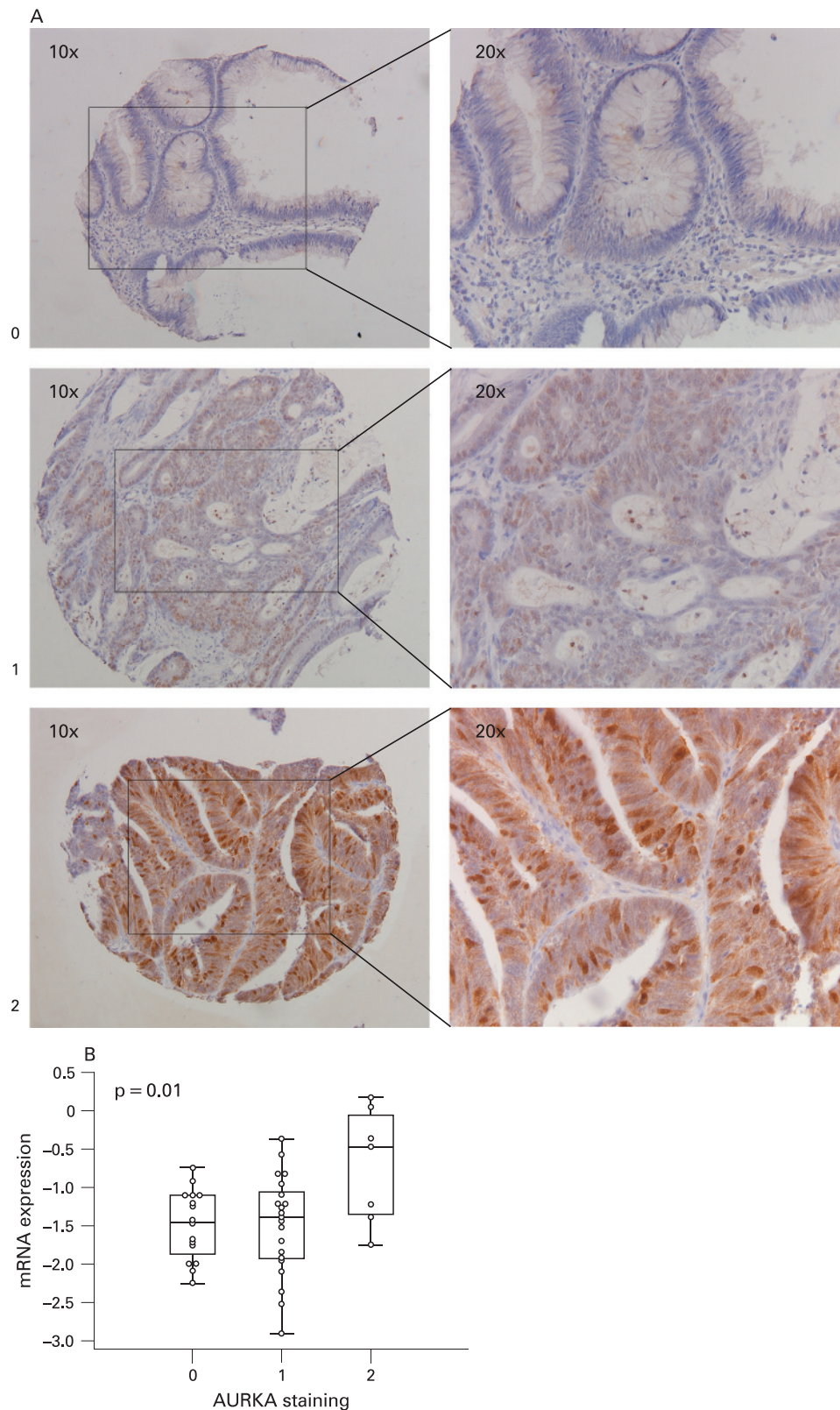
Little is known about the function of most of these genes. Some are transcription factors, like *TCFL5*,<sup>47</sup> or otherwise involved in transcriptional regulation, like *C20orf20*.<sup>48</sup> The product of *TH1L* is involved in regulation of A-Raf kinase.<sup>49</sup> *ADRM1* encodes a putative cell adhesion molecule that recently was shown to be a component of the 26S proteasome.<sup>50</sup> The product of *RNPC1* is predicted to bind to RNA, based on

**Table 4** Expression fold changes and range of expression levels (log<sub>2</sub> ratio) determined by expression microarray and by qRT-PCR, comparing either carcinomas vs adenomas (Ca/Ad) or tumours with 20q gain vs tumours without 20q gain (20q gain/non-20q gain)

Gene	Comparison	Microarray fold change	qRT-PCR fold change	Microarray expression range*	qRT-PCR expression range*
<i>C20orf24</i>	Ca/Ad	1.54	1.78	(−0.45 to 1.60)/(−0.71 to 0.71)	(1.84 to 6.08)/(−0.26 to 4.81)
	20q gain/non-20q gain	1.68	3.99	(−0.17 to 1.60)/(−0.71 to 0.37)	(−0.26 to 6.08)/(1.85 to 4.95)
<i>AURKA</i>	Ca/Ad	1.91	3.39	(−2.01 to 0.17)/(−2.26 to −1.11)	(−1.78 to 6.06)/(−0.64 to 3.72)
	20q gain/non-20q gain	1.55	4.53	(−2.11 to 0.17)/(−2.26 to −0.48)	(1.03 to 6.06)/(−1.78 to 3.99)
<i>RNPC1</i>	Ca/Ad	1.74	nd	(−1.61 to 1.22)/(−1.80 to −0.41)	nd
	20q gain/non-20q gain	1.58	nd	−1.71 to −1.22/−1.80 to −0.01	nd
<i>TH1L</i>	Ca/Ad	1.52	4.98	(−0.77 to 1.39)/(−1.06 to −0.15)	(−1.97 to 6.27)/(−3.57 to 3.72)
	20q gain/non-20q gain	1.59	6.4	(−0.59 to 1.39)/(−1.06 to 0.10)	(−3.57 to 6.27)/(−3.57 to 3.72)
<i>ADRM1</i>	Ca/Ad	1.45	1.46	(−0.62 to 0.79)/(−1.14 to 0.02)	(−0.30 to 5.58)/(−1.29 to 5.34)
	20q gain/non-20q gain	1.38	2.58	(−0.69 to 0.78)/(−1.14 to 0.36)	(−1.29 to 5.58)/(−0.30 to 5.34)
<i>C20orf20</i>	Ca/Ad	1.36	3.08	(−0.94 to 0.49)/(−1.31 to −0.59)	(−1.32 to 2.07)/(−2.79 to 0.14)
	20q gain/non-20q gain	1.34	3.57	(−0.89 to 0.49)/(−1.31 to −0.36)	(−1.16 to 2.06)/(−2.79 to 0.35)
<i>TCFL5</i>	Ca/Ad	2.2	3.54	(−2.14 to 0.83)/(−2.73 to −1.17)	(2.07 to 6.94)/(−1.28 to 4.21)
	20q gain/non-20q gain	2.02	3.54	(−2.31 to 0.83)/(−2.73 to −0.93)	(−1.28 to 6.94)/(1.99 to 4.41)

\*Log<sub>2</sub> ratio.

nd, not determined; qRT-PCR, quantitative reverse transcription-PCR.



**Figure 6** (A) Examples of AURKA protein expression in tissue microarray cores of an adenoma showing no expression (0), an adenocarcinoma showing weak expression (1) and an adenocarcinoma showing strong expression (2). (B) Combined box plot with dot plot of mRNA expression, determined by oligonucleotide microarrays (y-axis), of colorectal adenomas and carcinomas with a negative (0), weak (1) or strong (2) protein expression of AURKA on immunohistochemistry (x-axis).

sequence motifs, and *C20orf24* interacts with Rab-5, although its precise function is still unknown. *AURKA* has been well characterized and is involved in cell cycle regulation. It has been

shown to be amplified in CRC<sup>51</sup> and its overexpression induces centrosome amplification, aneuploidy and transformation in vitro.<sup>52</sup> Moreover, inhibiting *AURKA* by RNA interference led to

## Colorectal neoplasia

growth suppression of human pancreatic cancer cells.<sup>53</sup> Knocking down *TCFL5* resulted in suppression of the number of multicellular HT29 tumour spheroids, supporting its role in cancer development.<sup>54</sup>

In summary, we demonstrated the involvement of three SROs in the 20q amplicon in CRC and showed strong DNA copy number–mRNA expression associations for seven genes in these areas. In addition we demonstrated significant differences between colorectal adenomas and carcinomas at the DNA, mRNA and, for one of the genes, the protein level, supporting an important role as oncogenes in colorectal adenoma to carcinoma progression. Furthermore, we showed that expression levels of two out of the seven genes allowed adenomas to be discriminated from carcinomas with high accuracy. These genes therefore may serve both as highly specific biomarkers for CRC and as possible targets for pharmaceutical intervention in the development of CRC.

**Acknowledgements:** We thank the mapping core and map finishing groups of the Wellcome Trust Sanger Institute for initial BAC clone supply and verification. We would like to thank Anders Svensson for handling the array CGH database, Wessel van Wieringen for supervising the use of the software tool ACEit, Marianne Tijssen and Sjoerd J Vosse for helping with formatting all the raw data into the MIAME format, and Meike de Wit and Pien van Diemen for helping in the immunohistochemistry analysis.

**Funding:** This study was financially supported by the Dutch Cancer Society (KWF2002-2618).

**Competing interests:** None.

## REFERENCES

1. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990;**61**:759–67.
2. Muto T, Bussey HJ, Morson BC. The evolution of cancer of the colon and rectum. *Cancer* 1975;**36**:2251–70.
3. Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. *Nature* 1997;**386**:623–7.
4. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* 2004;**96**:261–8.
5. Edelmann L, Edelmann W. Loss of DNA mismatch repair function and cancer predisposition in the mouse: animal models for human hereditary nonpolyposis colorectal cancer. *Am J Med Genet C Semin Med Genet* 2004;**129**(1):91–9.
6. di Pietro M, Belver JS, Menigatti M, et al. Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers. *Gastroenterology* 2005;**129**:1047–59.
7. Hermesen M, Postma C, Baak J, et al. Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. *Gastroenterology* 2002;**123**:1109–19.
8. Weiss MM, Kuipers EJ, Postma C, et al. Genomic profiling of gastric cancer predicts lymph node status and survival. *Oncogene* 2003;**22**:1872–9.
9. Rajagopalan H, Nowak MA, Vogelstein B, et al. The significance of unstable chromosomes in colorectal cancer. *Nat Rev Cancer* 2003;**3**:695–701.
10. Ried T, Knutzen R, Steinbeck R, et al. Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Genes Chromosomes Cancer* 1996;**15**:234–45.
11. Meijer GA, Hermesen MA, Baak JP, et al. Progression from colorectal adenoma to carcinoma is associated with non-random chromosomal gains as detected by comparative genomic hybridisation. *J Clin Pathol* 1998;**51**:901–9.
12. Douglas EJ, Fiegler H, Rowan A, et al. Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. *Cancer Res* 2004;**64**:4817–25.
13. Nakao K, Mehta KR, Fridlyand J, et al. High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. *Carcinogenesis* 2004;**25**:1345–57.
14. De Angelis PM, Clausen OP, Schjølberg A, et al. Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their associations with tumour DNA ploidy, genotypes and phenotypes. *Br J Cancer* 1999;**80**:526–35.
15. Nessling M, Solinas-Toldo S, Wilgenbus KK, et al. Mapping of chromosomal imbalances in gastric adenocarcinoma revealed amplified protooncogenes MYCN, MET, WNT2, and ERBB2. *Genes Chromosomes Cancer* 1998;**23**:307–16.
16. Korn WM, Yasutake T, Kuo WL, et al. Chromosome arm 20q gains and other genomic alterations in colorectal cancer metastatic to liver, as analyzed by comparative genomic hybridization and fluorescence in situ hybridization. *Genes Chromosomes Cancer* 1999;**25**:82–90.
17. Kallioniemi A, Kallioniemi OP, Piper J, et al. Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc Natl Acad Sci USA* 1994;**91**:2156–60.
18. Tanner MM, Grenman S, Koul A, et al. Frequent amplification of chromosomal region 20q12–q13 in ovarian cancer. *Clin Cancer Res* 2000;**6**:1833–9.
19. Aust DE, Muders M, Kohler A, et al. Prognostic relevance of 20q13 gains in sporadic colorectal cancers: a FISH analysis. *Scand J Gastroenterol* 2004;**39**:766–72.
20. Witting SM, Snijders PJ, Meijer GA, et al. Increased gene copy numbers at chromosome 20q are frequent in both squamous cell carcinomas and adenocarcinomas of the cervix. *J Pathol* 2006;**209**:220–30.
21. Weiss MM, Snijders AM, Kuipers EJ, et al. Determination of amplicon boundaries at 20q13.2 in tissue samples of human gastric adenocarcinomas by high-resolution microarray comparative genomic hybridization. *J Pathol* 2003;**200**:320–6.
22. Albertson DG, Ylstra B, Segraves R, et al. Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nat Genet* 2000;**25**:144–6.
23. Postma C, Hermesen MA, Coffa J, et al. Chromosomal instability in flat adenomas and carcinomas of the colon. *J Pathol* 2005;**205**:514–21.
24. Weiss MM, Hermesen MA, Meijer GA, et al. Comparative genomic hybridisation. *Mol Pathol* 1999;**52**:243–51.
25. Carvalho B, Buffart TE, Reis RM, et al. Mixed gastric carcinomas show similar chromosomal aberrations in both their diffuse and glandular components. *Cell Oncol* 2006;**28**:283–94.
26. Muris JJ, Ylstra B, Cillessen SA, et al. Profiling of apoptosis genes allows for clinical stratification of primary nodal diffuse large B-cell lymphomas. *Br J Haematol* 2007;**136**:38–47.
27. Quackenbush J. Microarray data normalization and transformation. *Nat Genet* 2002;**32**(Suppl):496–501.
28. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002;**30**:207–10.
29. Jong K, Marchiori E, Meijer G, et al. Breakpoint identification and smoothing of array comparative genomic hybridization data. *Bioinformatics* 2004;**20**:3636–7.
30. Menten B, Pattyn F, De Preter K, et al. arrayCGHbase: an analysis platform for comparative genomic hybridization microarrays. *BMC Bioinformatics* 2005;**6**:124.
31. van de Wiel MA, Smeets SJ, Brakenhoff RH, et al. CGHMultiArray: exact P-values for multi-array comparative genomic hybridization data. *Bioinformatics* 2005;**21**:3193–4.
32. Diskin SJ, Eck T, Greshock J, et al. STAC: a method for testing the significance of DNA copy number aberrations across multiple array-CGH experiments. *Genome Res* 2006;**16**:1149–58.
33. van Wieringen WN, Belien JA, Vosse SJ, et al. ACE-it: a tool for genome-wide integration of gene dosage and RNA expression data. *Bioinformatics* 2006;**22**:1919–20.
34. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 2001;**25**:402–8.
35. Dydensborg AB, Herring E, Auclair J, et al. Normalizing genes for quantitative RT-PCR in differentiating human intestinal epithelial cells and adenocarcinomas of the colon. *Am J Physiol Gastrointest Liver Physiol* 2006;**290**:G1067–74.
36. Notterman DA, Alon U, Sierk AJ, et al. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res* 2001;**61**:3124–30.
37. Lin YM, Furukawa Y, Tsunoda T, et al. Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. *Oncogene* 2002;**21**:4120–8.
38. Pilarsky C, Wenzig M, Specht T, et al. Identification and validation of commonly overexpressed genes in solid tumors by comparison of microarray data. *Neoplasia* 2004;**6**:744–50.
39. Shimokawa T, Matsushima S, Tsunoda T, et al. Identification of TOMM34, which shows elevated expression in the majority of human colon cancers, as a novel drug target. *Int J Oncol* 2006;**29**:381–6.
40. Platzer P, Upender MB, Wilson K, et al. Silence of chromosomal amplifications in colon cancer. *Cancer Res* 2002;**62**:1134–8.
41. Tsafirir D, Bacold M, Selvanayagam Z, et al. Relationship of gene expression and chromosomal abnormalities in colorectal cancer. *Cancer Res* 2006;**66**:2129–37.
42. Marumoto T, Zhang D, Saya H. Aurora-A—a guardian of poles. *Nat Rev Cancer* 2005;**5**:42–50.
43. Carter SL, Eklund AC, Kohane IS, et al. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 2006;**38**:1043–8.
44. Hidaka S, Yasutake T, Takeshita H, et al. Differences in 20q13.2 copy number between colorectal cancers with and without liver metastasis. *Clin Cancer Res* 2000;**6**:2712–7.
45. Rooney PH, Boonsong A, McFadyen MC, et al. The candidate oncogene ZNF217 is frequently amplified in colon cancer. *J Pathol* 2004;**204**:282–8.
46. Brinkmann U, Gallo M, Polymeropoulos MH, et al. The human CAS (cellular apoptosis susceptibility) gene mapping on chromosome 20q13 is amplified in BT474 breast cancer cells and part of aberrant chromosomes in breast and colon cancer cell lines. *Genome Res* 1996;**6**:187–94.
47. Siep M, Sleddens-Linkels E, Mulders S, et al. Basic helix–loop–helix transcription factor Tcf15 interacts with the Calmegin gene promoter in mouse spermatogenesis. *Nucleic Acids Res* 2004;**32**:6425–36.
48. Cai Y, Jin J, Tomomori-Sato C, et al. Identification of new subunits of the multiprotein mammalian TRRAP/TIP60-containing histone acetyltransferase complex. *J Biol Chem* 2003;**278**:42733–6.



49. **Liu W**, Shen X, Yang Y, *et al*. Trihydrophobin 1 is a new negative regulator of A-Raf kinase. *J Biol Chem* 2004;**279**:10167–75.
50. **Jorgensen JP**, Lauridsen AM, Kristensen P, *et al*. Adrm1, a putative cell adhesion regulating protein, is a novel proteasome-associated factor. *J Mol Biol* 2006;**360**:1043–52.
51. **Bischoff JR**, Anderson L, Zhu Y, *et al*. A homologue of Drosophila aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J* 1998;**17**:3052–65.
52. **Zhou H**, Kuang J, Zhong L, *et al*. Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat Genet* 1998;**20**:189–93.
53. **Hata T**, Furukawa T, Sunamura M, *et al*. RNA interference targeting aurora kinase A suppresses tumor growth and enhances the taxane chemosensitivity in human pancreatic cancer cells. *Cancer Res* 2005;**65**:2899–905.
54. **Dardousis K**, Voolstra C, Roengvoraphoj M, *et al*. Identification of differentially expressed genes involved in the formation of multicellular tumor spheroids by HT-29 colon carcinoma cells. *Mol Ther* 2007;**15**:94–102.

## Editor's quiz: GI snapshot

### ANSWER

From the question on page 24

Colonoscopy revealed a localised area of mucosa inflammation with a suspected fistulous tract opening at the colonic hepatic flexure. Abdominal computed tomography (CT) and ultrasound (US) showed the presence of a large gallstone occupying the entire gallbladder and diffuse pneumobilia that raised the suspicion of a biliary–enteric fistula. Subsequent endoscopic retrograde cholangiopancreatography showed an initial compression of the common bile duct due to a large impacted gallstone, with minimal dilatation of the intrahepatic ducts. It also showed a cholecysto-colic fistula from which contrast entered the colon delineating the right colonic haustration (fig 1A,B, below).

A diagnosis of large cholecysto-colic fistula was made with associated Mirizzi syndrome type I.<sup>1</sup> Biliary–enteric fistulas are rare disorders usually developing insidiously and an association with gallstones is always present. The most common type of biliary–enteric fistula is cholecystoduodenal (75%); cholecysto-colic is next common (15–20%) with a variety of other types (ie, cholecystogastric) being less frequent (5–15%).<sup>2</sup>

Clinical manifestations of biliary–enteric fistulas are often not specific including recurrent cholangitis, right upper quadrant pain, jaundice, persistently abnormal liver biochemistry and, rarely, gallstone ileus. It is remarkable that the predominant symptoms in our patient were chronic diarrhoea and weight loss. The main mechanism underlying these complaints is the malabsorption of bile acids: since most of bile acids by-passed the small intestine through the cholecysto-colonic fistula this resulted in a diminished fat absorption; in addition, the unusual high amount of bile



**Figure 2** Large impacted stone in the gallbladder retrieved at surgery.

acids passing into the colon interfered with water absorption, causing bile acid diarrhoea.

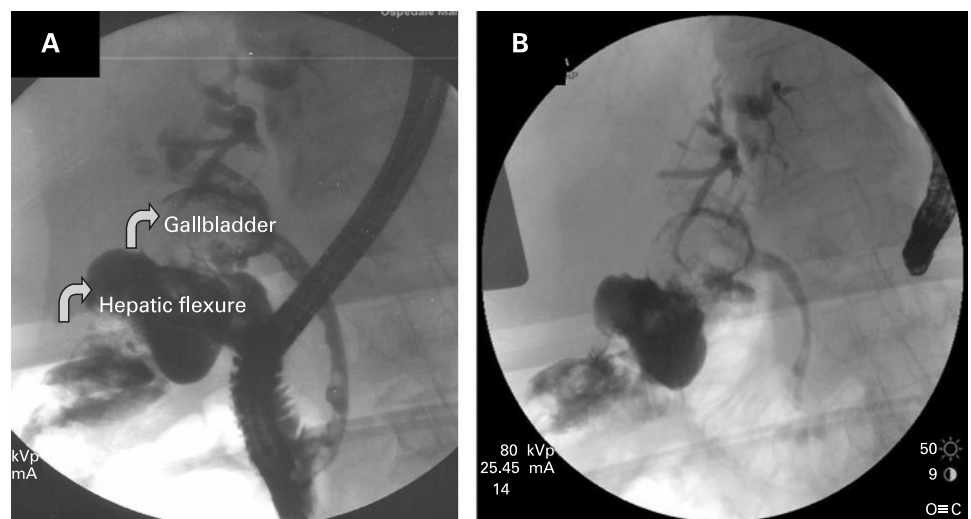
The patient underwent surgery and fistulectomy with cholecystectomy (fig 2) and Roux-en-Y biliary–enteric anastomosis resulted in prompt relief of symptoms.

*Gut* 2009;**58**:89. doi:10.1136/gut.2008.158667a

### REFERENCES

1. **Johnson LW**, Sehon JK, Lee WC, *et al*. Mirizzi's syndrome: experience from a multi-institutional review. *Am J Surg* 2001;**67**:11–4.
2. **Inal M**, Oguz M, Aksungur E, *et al*. Biliary–enteric fistulas: report of five cases and review of the literature. *Eur Radiol* 1999;**9**:1145–51.

**Figure 1** Cholangiograms showing cholecysto-colic fistula and minimal dilatation of intrahepatic ducts.







# Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression

B Carvalho, C Postma, S Mongera, et al.

*Gut* 2009 58: 79-89 originally published online October 1, 2008

doi: 10.1136/gut.2007.143065

---

Updated information and services can be found at:

<http://gut.bmj.com/content/58/1/79.full.html>

---

*These include:*

## Data Supplement

"web only appendix"

<http://gut.bmj.com/content/suppl/2008/12/05/58.1.79.DC1.html>

## References

This article cites 54 articles, 22 of which can be accessed free at:

<http://gut.bmj.com/content/58/1/79.full.html#ref-list-1>

Article cited in:

<http://gut.bmj.com/content/58/1/79.full.html#related-urls>

## Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

---

## Topic Collections

Articles on similar topics can be found in the following collections

[Colon cancer](#) (2932 articles)

---

## Notes

---

To request permissions go to:

<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:

<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:

<http://group.bmj.com/subscribe/>